

УДК 534.4: 621.391

О.М. Карпов, К. Г. Чебров

*Дніпропетровський національний університет ім. Олеся Гончара***РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМ СИНТЕЗУ
ЯКІСНОГО МОВЛЕННЯ ЗА ПРАВИЛАМИ**

Наведено опис практичної реалізації алгоритму синтезу мовленевих сигналів.

Ключові слова: *алгоритм, синтез мовлення, формантний метод.*

Приведено описание практической реализации алгоритма синтеза речевых сигналов.

Ключевые слова: *алгоритм, синтез речи, формантный метод.*

In the article were the description of the practical realization of the algorithm of the synthesis of the speech signal considered.

Key words: *algorithm, Speech synthesis, formant synthesis method.*

Вступ. Усі способи синтезу мови можна поділити на групи [1–4].

1. Параметричний синтез мови є кінцевою операцією у вокодерних системах, де мовний сигнал представляється набором невеликого числа параметрів, що безупинно змінюються. Параметричний синтез доцільно застосовувати в тих випадках, коли набір повідомлень обмежений і змінюється не надто часто. Перевагою такого способу є можливість записати мову для будь-якої мови і будь-якого диктора. Якість параметричного синтезу може бути дуже високою (залежно від ступеня стиску інформації за параметричним поданням). Однак параметричний синтез не може застосовуватися для довільних, заздалегідь не заданих повідомлень.

2. Компіляційний синтез зводиться до складання повідомлення з попередньо записаного словника вихідних елементів синтезу. Очевидно, що зміст синтезованих повідомлень фіксується обсягом словника. Основна проблема в компілятивному синтезі – обсяги пам'яті для зберігання словника. У цьому зв'язку використовуються різноманітні методи стиску/кодування мовного сигналу.

3. Повний синтез мови за правилами (або синтез за друкованим текстом) забезпечує керування всіма параметрами мовного сигналу і, таким чином, може генерувати мову по заздалегідь невідомому тексту.

У цьому випадку параметри, отримані при аналізі мовного сигналу, зберігаються у пам'яті так само, як і правила з'єднання звуків у слова і фрази. Синтез реалізується шляхом моделювання акустичних процесів мовного тракту, застосуванням аналогової або цифрової техніки. Причому в процесі синтезування значення параметрів і правила з'єднання фонем вводять послідовно через певний часовий інтервал, наприклад 5–10 мс.

Виділяються два підходи. Перший – спрямований на побудову моделі мовної системи людини, він відомий під назвою артикуляторний синтез. Другий підхід – формантний синтез за правилами. Розбірливість і натуральність таких синтезаторів може бути доведена до величин, порівняних з характеристиками природної мови.

Загальна частина. Реальна побудова синтезатора утримує деяку сукупність правил усіх типів синтезаторів з обліком тривалості частоти, інтенсивності і варіативності (не стаціонарності) параметричних представлень мовних одиниць.

Спектр мови $Y(\omega)$ можна зобразити як добуток частотної функції джерела (джерела збудження) $X(\omega)$ та перехідної функції мовного тракту $H(\omega)$ (1):

$$Y(\omega) = X(\omega) H(\omega) \quad (1)$$

За часовою залежністю мовний сигнал буде представлятиметься відповідною згортою (5.2) (для нерекурсивного фільтра):

$$y(t) = \int_0^{\infty} x(t)h(t - \tau)d\tau \quad (2)$$

Для одержання необхідної F-картини фонемі може бути використаним банк фільтрів. Для одержання спектру фонемі /А/, яка була вимовлена автором даної роботи в останньому ненаголошеному складі слова «дама», який зображено на рис. 1, може використовуватися банк з п'яти фільтрів (відповідно до кількості формант).

Нехай зображений спектр фонемі має формуватись на базі спектра джерела збудження, котрий має вигляд як на рис. 2.

Для цього формують банк фільтрів, параметрами котрих задають вид формант, приведених на рис. 3.

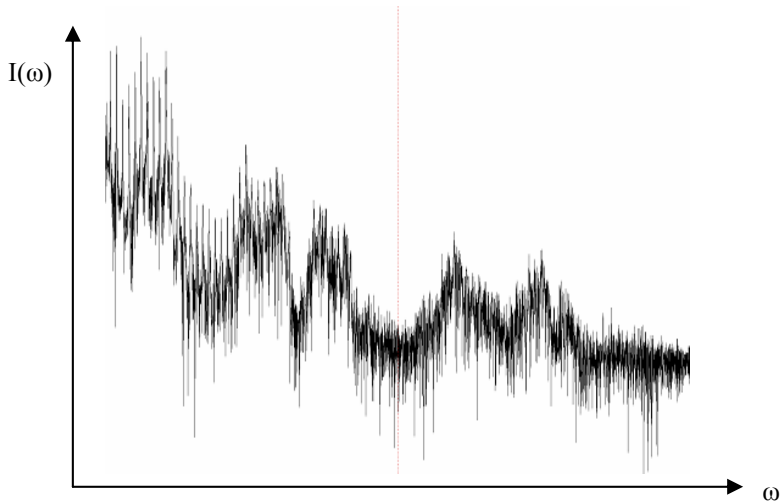


Рис. 1. Спектр фонем /А/

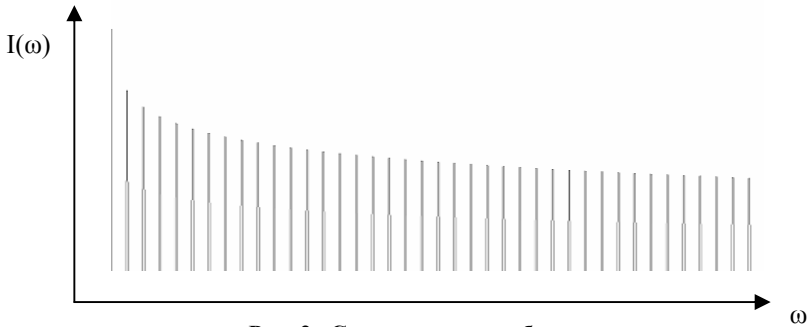


Рис. 2. Спектр джерела збудження

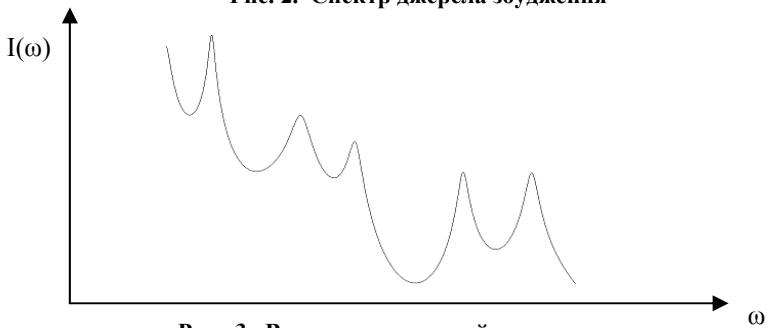


Рис. 3. Вид спектру, котрий породжується передатною функцією мовного тракту у вказаному випадку

Перейдемо до опису організації обчислювального процесу на базі наведених викладок. Програма має містити гнучкий інтерфейс. Вона реалізує синтез мовного сигналу за текстом на основі методу форматного синтезу з використанням можливостей компіляційного.

Вхідними даними є текстові файли, що містять інформацію про траєкторії зміни F-картини фонем, траєкторії інтенсивностей, значення основних параметрів, які можуть варіюватись, wav-файли, базовий для синтезу текст. Вихідними даними є синтезований (відновлений за текстом) звук, котрий зберігається у wav-файл.

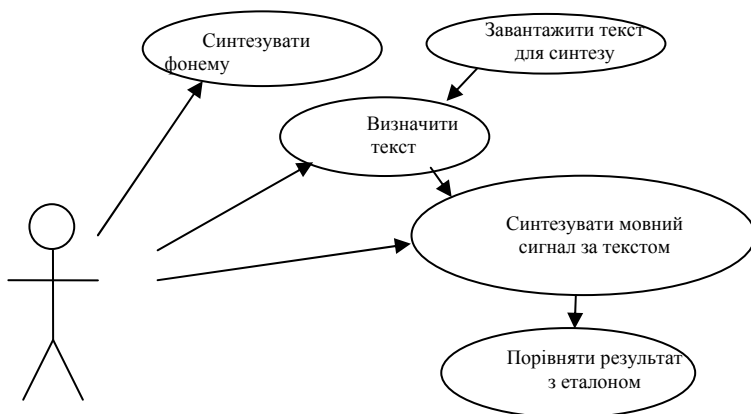


Рис. 4. Use-case діаграма програми

Програму було реалізовано засобами середовища MATLAB R2010a.

До програми входять такі модулі:

`amplMod.m` – утримує службові функції для реалізації амплітудної модуляції сигналу;

`applyFilter.m` – слугує для таких задач як генерація сигналу голосового джерела збудження, побудова результуючого спектру вокалізованої фонем за заданими параметрами;

`fricative.m` – слугує для таких задач як генерація сигналу шумового джерела збудження, побудова результуючого спектру фрікативних фонем за заданими параметрами;

`gui.m` – головний модуль, котрий збирає до сукупності та організовує взаємодію функціональних модулів з користувачем з програмою;

nValsSlide.m – містить допоміжні функції для забезпечення неперервного (не скачковидного) способу зміни параметрів процесу синтезу;

Synth.m – головний модуль, котрий поєднує функціональні підмодулі синтезатора;

singleFilter.m – містить відносно низькорівневі функції для завдання однієї форманти.

Можливості програми та шлях взаємодії користувача з нею можна представити на рис. 4 за допомогою use-case діаграми.

Продемонструємо роботу програми. Згенеруємо слово «антивирусные». При цьому програма має вигляд як на рис. 5.

На рис. 6 и рис.7 наведено вигляд вікон з внутрішньою проміжною інформацією у графічному вигляді, що використовується при синтезі вище згаданого слова – траєкторії зміни F-картини та інтенсивності відповідно.

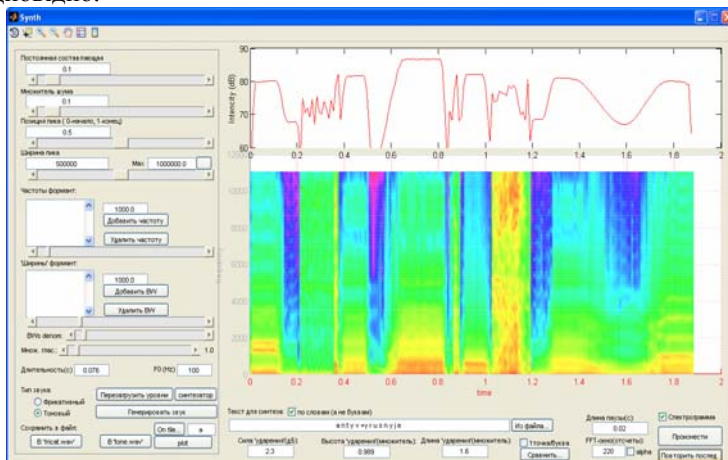


Рис. 5. Вигляд головного вікна програми після синтезу тексту

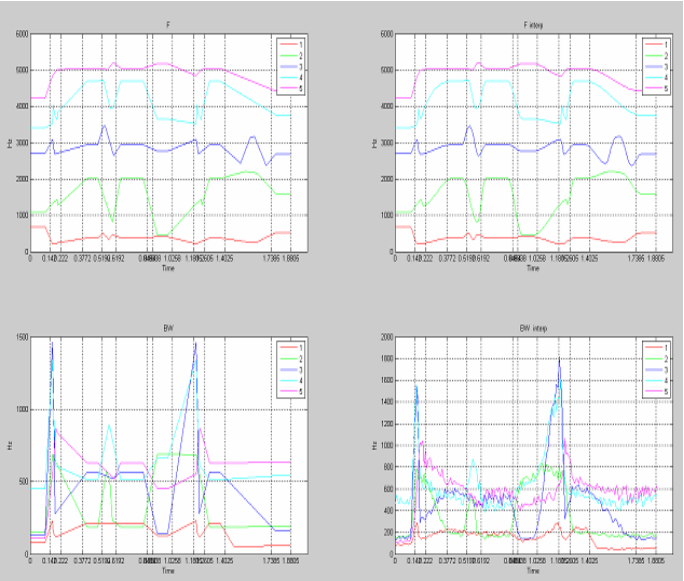


Рис. 6. Форматні траєкторії

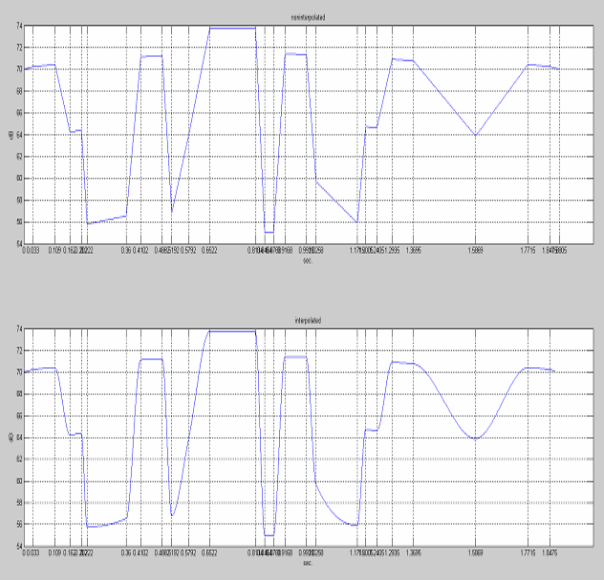


Рис. 7. Траєкторії інтенсивності

Під час проведення тестів програма показала коректну поведінку, відсутність помилок часу виконання.

Висновки. Зроблено програмний продукт для синтезу мовного сигналу з використанням методів формантного та компіляційного синтезу. Ця розробка відноситься до області надання автоматичним/автоматизованим системам мовного інтерфейсу за напрямком «машина–людина». Для реалізації поставленої задачі використовувалось середовище MATLAB R2010a.

Було зроблено огляд досягнень у галузі синтезу мовного сигналу, який показав, що провідні організації у всьому світі цікавляться та займаються дослідженнями в цьому напрямку. Але результати роботи існуючих систем не є повною мірою задовільними. Це обумовлено складністю процесу синтезу, зокрема складністю реалізації відображення перехідних процесів, нелінійностей та інших.

Проаналізувавши результати роботи програми, можна сказати, що, звичайно, якість синтезованої мови не дозволяє казати про передові показники, з чого слідує необхідність подальшого розвитку системи та використання додаткових методів, залучення механізмів артикуляційного синтезу. Під час проведення тестів програма показала коректну поведінку, відсутність помилок часу виконання.

Бібліографічні посилання

1. **Фант Г.** Акустическая теория речеробразования / Г. Фант. – М., 1964 – 284 с.
2. **Фланаган Д.** Анализ, синтез и восприятие речи / Д. Фланаган. – М., 1968 – 296 с.
3. **Сорокин В.Н.** Синтез речи / В.Н. Сорокин. – М., 1992. – 392 с.
4. **Лобанов Б.М.** Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск, 2008. – 339 с.

Надійшла до редколегії 15.06.2012